ABSTRACT OF THE DISCLOSURE

A method of identifying electronic files comprising the steps of identifying a beginning of the content within a file being transmitted through a network, generating a tag based on content of the file, and comparing the tag to other tags in a database of tags to measure similarity between the tag and the other tags.